DOCUMENT RESUME

ED 447 165 TM 032 007

AUTHOR	Canning-Wilson, Christine
TITLE	Marking English Writing Scripts: A Study Initiative between Blind Marking and Transparent Scoring Practices.
PUB DATE	2000-00-00
NOTE	17p.
PUB TYPE	Reports - Research (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Computer Uses in Education; Criterion Referenced Tests; *English (Second Language); Foreign Countries; *Grades (Scholastic); Higher Education; *Language Tests; Norm Referenced Tests; *Scoring; Student Evaluation; Test Use
IDENTIFIERS	*United Arab Emirates; United Arab Emirates University

ABSTRACT

The differences between the testing practices at the University General Requirements Unit (UGRU) of United Arab Emirates University and the Higher Colleges of Technology (HCT) were studied, focusing on the differences in styles of grading written English examinations. The UGRU always uses norm- referenced testing in their 1-year foundations program, while the HCT always uses criterion-referenced testing. Research shows that the HCT has more reliable scoring methods and testing practices than the UGRU. The successful scoring of more than 9,000 graded scripts in June 2000 by the HCT supports this finding. The paper also discusses why the use of blind markers and software packages by the HCT and Academic Services makes their grading practices fair and reliable examples for all institutions in the United Arab Emirates to follow. (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 447 165

Marking English Writing Scripts: A Study Initiative Between Blind Marking and Transparent Scoring Practices

By: **Christine Canning- Wilson**
Center for Excellence in Applied Research and Training (CERT), HCT – Abu Dhabi

Abstract:
This paper will look at the differences between the testing practices at the University General Requirements Unit (UGRU) at United Arab Emirates University (UAEU) and the Higher Colleges of Technology (HCT). The paper focuses on the differences between the 2 institutions' styles of grading English written exams. Research will show that the HCT has more reliable scoring methods and testing practices than the UGRU. The paper documents the successful scoring of 9,000 graded scripts in June 2000 by the HCT. Additionally, this paper will look at why the HCT's and Academic Services' use of blind markers and use of software packages make grading of exam scripts a fair and reliable practice for all institutions in the UAE to model.

1.0 Literature Review

Many programs, with high-stakes exams administered on large or small-scale use either norm-reference testing or criterion-based testing practices. Norm-reference tests derive their pass mark from the performance of the student. It is not predetermined. Criterion-reference tests are written to reflect a specific curriculum. Criterion- reference tests may also use criteria separate from outside the given curriculum and/or student population to determine proficiency. The UGRU uses norm-reference testing in their one-year foundations program. The HCT always uses criterion-reference testing, to measure student performance, proficiency and mastery of the English language.

The forerunning experts in testing, Alderson and Wall, theorize a washback hypothesis between assessment and its relationship to teaching and learning. In their cornerstone article, Alderson and Wall (1993) state:

- A test will influence teaching.
- A test will influence learning.
- A test will influence what teachers teach.
- A test will influence what learners learn.
- A test will influence how a teacher teaches.
- A test will influence how a learner learns.
- A test will influence the rate and sequence of teaching.
- A test will influence the rate and sequence of learning.
- A test will influence the degree and depth of teaching.

TM032007

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Christine Canning-Wilson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

- A test will influence the degree and depth of learning.
- A test will influence attitudes to the content, method, etc. teaching and learning.
- Tests that have important consequences will have washback.
- Tests that do not have important consequences will not have washback.
- Tests will have washback on all learners and teachers.
- Tests will have washback effects for some learners and not for others.

Additional factors beyond those stated by Anderson and Wall, affect learning, teaching and testing. Other influences and variables must be entered into the testing process in order to understand how a test impacts learning, teaching, curriculum and/or administrative practices. I believe these practices and beliefs include, but are not limited to:

- How a test influences washback/feedback for teaching and learning is dependent on each individual program and its curriculum.
- A test is only a test. It is not necessarily the reflection of one's ability to understand all information. Scores reflect the results to a set of predisposed questions.
- The type of test written may reflect internal standards of a set curriculum or program.
- A set of test questions may be reliable and valid sources to create an assessment instrument, even if the grading of the instrument is invalid or unreliable.
- Criteria used to classify a test as the piloting or non-piloting of test question influences a valid and reliable instrument.
- Administrative constraints time limitations and course content influence teaching.
- Administrative constraints time limitations and course content influence learning.
- Grading criteria influence teaching practices as well as learning practices.
- Implementing grading internationally recognized criteria influences standards.
- Testing influences the impact of grading practices and policies both in and out of the examination environment.
- Grading influences the impact of testing.
- Training of teachers to grade consistently with standards impacts the quality of test results and exam preparation.
- The lack of training to grade consistently with standards impacts the quality of test results and exam preparation.
- Testing without proper statistical analysis can lead to incomplete feedback.
- Descriptive statistics can be manipulated and may not offer authentic feedback into the teaching, learning and testing process.
- Changes in policies, procedures or styles of grading on a test will influence scores. Scores that do not reflect consistency cannot give proper feedback to a program.
- Task type will influence learning, teaching and testing.
- Grading to standard will warrant better accuracy into the goals of teaching and learning.
- Standardized grading practices will positively influence and impact testing, teaching and learning.

- Non- standardized grading practices will negatively influence and impact testing, teaching and learning.

2.0 Program Overview

The University General Requirements Unit is a one-year foundations program. The UGRU curriculum serves as a transitional phase of instruction for students leaving high school and entering their respective faculties. The UGRU provides up to three semester-long levels of instruction. Each level (English 1, English 2, and English 3) offers approximately nine hours per week of basic and intermediate instruction in English. These courses carry credit and lead to English-for-Special Purpose (ESP) classes in a separate ESP Program. UGRU English 2 and 3 courses have dual tracks, one for students whose major requires extensive amounts of English (ESL track), and a less intensive one for students studying in the EFL track. For example, most science majors are in the ESL track; whereas, most Education and Arts majors are in the EFL tracks. As of Spring 2000, the UGRU students were assessed with a 10% teacher grade, 30% mid-term grade and a 60% final exam grade based on in-house teacher produced examinations overseen by the testing committee.

In comparison to the UGRU program, the Higher Colleges of Technology offers solid programs, which can lead to a Certificate, Diploma, Higher Diploma or Bachelor's Degree using English as a medium of instruction. The Colleges use both college based assessment examinations and international benchmark exams such as the PET and the IELTS to measure students' proficiency.

Both institutions are dedicated to improving the English language skills of Emirati nationals. Their commitments to furthering language abilities of young nationals can be found in their in-depth mission statements. The UGRU On-Line Mission Statement is stated verbatim as follows:

> "The University General Requirements Unit (UGRU) seeks to equip all incoming students with basic skills and scientific tools necessary for creativity and taking initiatives for the proper acquisition of knowledge. UGRU strives to develop the educational process into an active and interactive learning environment." (http://www.ugru.uaeu.ac.ae/)

It further states that "…Because students will interact with English to different degrees and in varying contexts, individualization *must be built into proficiency-oriented curricula*." As of the time of the study, no official use of proficiency guidelines standardized tests (PET/IELTS) or formal speaking courses were in practice.

The Mission Statement and goals of the Higher Colleges of Technology (2000) differ from its sister institution. The HCT Handbook and On-line Handbook 2000 states that:

> *"The Higher Colleges of Technology (HCT) are dedicated to the delivery of technical and professional programs of the highest quality to the citizens of the*

United Arab Emirates (UAE). Graduates of the colleges will have the linguistic ability to function effectively in an international environment; the technical skills to operate in an increasingly complex technological world; the intellectual capacity to adapt to constant change, and the leadership potential to make the fullest possible contribution to the development of the community for the good of all its people."
(HCT Handbook 2000 ref:http://www.hct.ac.ae/general_info/frame.htm)

The Aims of the English Program according to the HCT Handbook are stated verbatim:

> *Aims (ref:http://imtcsamba.hct.ac.ae/gened/english/MissionStatement.html)*
>
> *"English Language teaching in the HCT supports the objectives of the HCT mission statement.*
>
> *Graduates of the HCT will have the linguistic ability to function effectively in an international environment. This statement recognizes that English is increasingly the medium of the workplace and the language of modern international business, technical and educational communities. HCT English Language Curriculum prepares students to meet these demands by teaching the relevant transferable language skills to function in whatever employment they may choose.*
>
> *The HCT is an English medium institution; therefore, it is essential to equip students with the English Language skills sufficient to follow their course of study. This includes the relevant study skills and cognitive skills training to enable the student to process, organize and present their ideas in an academic or vocational context.*
>
> *English Language teaching plays a major role in students' personal development.*
>
> *It provides the language skills to engage in social relations, first, in the college environment and later in wider society and the workplace. It gives students sufficient language skills to deal with their everyday needs in English. English Language teaching also focuses on the development of literacy skills which enable students to better understand written information, and to express themselves more effectively in writing. As students are exposed, through English, to a wide range of information and ideas they are encouraged to discuss and relate their new knowledge of the world to themselves.*

HCT students learn to access and manage information in English for work and leisure purposes, and to use new technologies confidently. This enables HCT students to become independent life long learners. The language, technology, and thinking skills they acquire at HCT equip them to make a valuable contribution to their society, through their roles in the workplace, in the community and in the home."

The Curriculum (Ref: http://imtcsamba.hct.ac.ae/gened/english/MissionStatement.html)

"The role of the English curriculum is therefore a complex and crucial one, addressing far wider needs than the purely linguistic.

The English Language curriculum is designed to meet the aims of the mission statement. The starting point of the curriculum is the assessment of student needs and employer requirements in relation to those aims. Student English language needs are assessed by administering the English Diagnostic Assessment on entry to the HCT. Employer English language needs are monitored by surveys and feedback from employers. The process results in a skills based curriculum where student language proficiency is described in terms of performance outcomes.

Graduates of the Certificate program can initiate and sustain conversations on everyday topics. They are able to process and exchange simple written information in routine everyday personal and workplace situations.

Graduates of the Diploma program can communicate orally on a range of familiar vocational and social topics. They are able to process and exchange straightforward written information in a wider range of familiar personal and workplace situations.

Graduates of the Higher Diploma program can interact orally with flexibility in a varied range of situations. They can process and exchange written information and ideas with some degree of complexity in a variety of personal, academic and workplace situations.

The English syllabus provides students with lexical, functional, grammatical and micro skill inputs to achieve the incremental performance outcomes."

Methodology (Ref:http://imtcsamba.hct.ac.ae/gened/english/MissionStatement.html)

"The HCT English language teaching learning methodology focuses

on communicative use of language skills for meaningful exchanges in realistic situations.

Classroom approaches will move students from their previous teacher-dependent learning habits, which have emphasized surface memorization techniques, towards a more independent and self evaluative style where students will develop the ability to comprehend, process, analyze and synthesize information in English. As students' cognitive and language skills are developed, the content and level of materials will move from basic information based tasks towards more complex processing and expression of ideas.

Students learn to develop their skills through appropriate strategies including role-play, pair and group work. Students will also use computer software for interactive language learning activities, for self study, as well as using the Internet and word-processing packages as regular part classroom activities. Self-study habits are encouraged through the use of Independent Learning Centers and extensive reading through using materials in LRC's.

Reading skills and strategies are taught explicitly and extended from basic level through intensive focus on micro skills using a variety of authentic text types and instructor adapted materials. Writing skills are similarly taught from a basic level through materials which address the particular needs of Arab students, moving at higher levels towards the structured expression of ideas. Syllabus inputs are taught using communicative style international texts selected by the course teams. Frequent recycling of these inputs is undertaken to consolidate learning.

Skills are taught and practiced in an integrated fashion, so that students can transfer information and ideas from one mode to another, e.g. by reading and reporting findings in writing.

Students are required to transfer their learned skills to a variety of contexts that mirror the reality of the workplace and the wider world. To enhance this process students are exposed to a wide range of contexts and materials. These contexts will include workplace-related materials associated with the students' program areas."

Assessment:
(Assessment Handbook - Ref: http://imtcsamba.hct.ac.ae/stand_eval/assessment/)

" Assessment activities mirror teaching and learning activities being

communicative and task based in nature.

Assessment consists of in-course formative and summative assessment of objectives, followed by end of course proficiency-style checks using common specifications. Performance objectives are directly assessed using performance-based criteria. Performance descriptors based on international nine-point scales are used to assess proficiency in speaking and writing. At key points four-skills proficiency assessments are constructed and administered at a system-wide level to ensure consistency across colleges. In addition, Diploma and Higher Diploma programs also have an international proficiency benchmark to further validate the standards of in-house assessment."

Structure (Ref: http://imtcsamba.hct.ac.ae/gened/english/MissionStatement.html)

" The delivery of English Language mission objectives is the responsibility of the English Subject Team, who report, through Academic Services, to the Academic Council. The systemwide delivery of curriculum course content is the responsibility of system course teams and system Course Coordinators, who in turn report to the English Subject Team. Within each college the relevant teaching Supervisor is responsible for the day to day implementation of the curriculum.

A constant cycle of feedback, evaluation and review is built into the process at all levels to ensure quality is achieved."

It is clear that the mission of the Higher Colleges of Technology is to improve teaching excellence and to maximize opportunities, which keeps learning on the cutting edge of technological and pedagogical advances. Unlike UAEU students, HCT students are required to take English courses through out their entire program of study until graduation. Teachers are vigorously trained to comply with the high standards set forth by the Higher Colleges of Technology and Academic Services. The Higher Colleges of Technology (HCT) are dedicated to the delivery of technical and professional programs of the highest quality to the citizens of the United Arab Emirates (UAE).

3.0 Testing Programs

There are other fundamental differences between the UGRU and the HCT programs. Firstly, the Higher Colleges of Technology offers and tests a speaking component. To date the University General Requirements Unit at UAE University does not formally teach or test speaking. Secondly, the Higher Colleges uses standardized tests. These regularly administered and highly recognizable examinations allow HCT students to be benchmarked by international standards. UGRU courses at UAE University do not use standardized exams that have been piloted and validated at the end

of their courses. Instead, level heads and teachers create exams based on the materials found in the books using similar exercises. Section 3 of the UGRU Mission Statement On-line Program File states, "Students are first assessed through a placement test. The midterm and final examinations are developed by the Testing Committee working collaboratively with all the teachers."

The UGRU exams are produced in-house, and are not calibrated against international benchmark tests. The test is written to the level of the student and the course materials, this practice often result in an approximate 70/30-pass/fail rate on examinations. This is a significant achievement in comparison to the 1993 BUEC/UGRU goal to aim for a proposed 60% pass rate on in-house tests. To obtain these new results during the seven-year period, self-assessment reviews and new testing practices had to be implemented. As the HCT's Academic Services support to student assessments are standardized and centralized, the ACS was not faced with the same types of challenges as the UGRU.

Because the UGRU does not follow all of the same basic testing practices employed by the HCT, UGRU exams, over the years, have been known to have more than one answer to a question, typing errors and to have had irregular pass/fail rates where upper-administration has had to call the testing process into question.

An additional challenge the UGRU's testing committee has faced is the need to produce criterion exams, which are written to the level of the students. These in-house exams, because of previous mentioned constraints make it next to impossible for the score results to be benchmarked at international standards. Furthermore, unlike standardized exams, which are piloted and have a regular system for validating questions, the in-house produced tests at UGRU potentially run the risk of problems with testing issues such as validity. The last challenge faced by the UGRU, and to be fair most testing programs worldwide is test security. As stated before, because of the number of teachers on curriculum and testing teams, who have potential access to exams, there still remains a potential risk as far as security issues are concerned. Again, because the HCT's Academic Services are standardized and centralized it is not faced with the same types of challenges.

4.0 Grading Criteria for Marking Scripts:

Data examined shows that the HCT has a more reliable system of examination writing, vetting and scoring than the UGRU. Computerized statistical reports show that grading is more accurate and objective at the HCT. Practices for grading are based on their own HCT in-house writing bands, which are calibrated with the internationally recognized ESU Framework.

Empirical evidence will show that the Higher Colleges of Technology's system for banding is a more reliable way of scoring than that of the current UGRU practice of in-house designed calibration.

Each semester, the ACS at HCT offers new and veteran faculty on-going training seminars in grading writing scripts. Furthermore, it trains its teachers to meet international grading requirements to ensure quality and equity amongst markers. As stated previously, the HCT in-house banding system is based on the ESU Framework. In other areas of evaluation and assessment, the HCT sponsors international and prestigious

testing conferences such as the Current Trends in English Language Testing (CTELT). It has hired testing specialists who are internationally recognized. It offers its employees handbooks and on-line self-access manuals for reference. Most recently, the HCT encouraged the development of teacher and student sites for testing practice with its generous support from QTL Grants. It would be more than fair and accurate to state that the HCT is a forerunner in its commitment to furthering good testing practices.

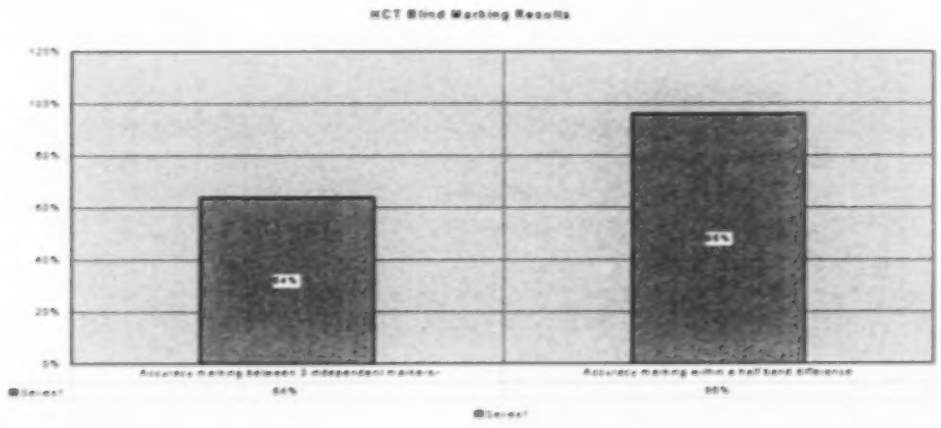
As a result of the in place standards, the HCT students are the benefactors of quality control. Learners are regularly graded and assessed based on internally created and simulated international benchmarks as well as by formal standardized examinations of proficiency. The HCT's stringent use of academic requirements such as the UCLES/IDP, International English Language Testing System (IELTS) and the UCLES Preliminary English Test (PET), allow the HCT learners to be eligible to actively compete in a global world using the English language. HCT students are regularly encouraged to apply and learn the language for use in the workplace and to benchmark proficiency against other language learners around the world. Moreover, passing these prestigious international examinations is required for students to graduate from an HCT Program.

In comparison, the UGRU does not incorporate standardized exams at the end of its courses to grade a student's proficiency. As stated earlier in the UGRU Mission Statement, they still must try to build proficiency based-curricula. Therefore, they have no reason to test for proficiency in the language at the current time, since it is not taught in their program. As a result, they rely solely on their teacher-generated tests.

HCT college-based exams are graded by an in-house designed banding system based on the ESU Framework for banding. Bands range from 1-10 with quarter band intervals to distinguish ability and score differentials. Formal exterior assessment of University of Cambridge Local Examination Syndicate (UCLES) exams such as IELTS and PET use the established standardized criteria.

A review of an approximate 9,000 written script exams scored in June 2000 revealed a 64% marking accuracy in same score grading between three independent markers using blind marking practices (see figure 1a). Software statistical practices showed a 96% marking accuracy rate within a half band radius amongst three independent markers using blind marking practices. These results are phenomenal and the HCT practices should serve as a model for other institutions in the Arab world.

Figure 1a:



The UGRU writing calibration practices differ greatly from those in use at the HCT. The calibration criteria to determine student grades changes semester to semester and sometimes exam to exam. In hand, I have 5 different BUEC/UGRU calibration sheets for grading, but for the purposes of this paper, I will use the calibration sheet devised for the Spring 1998 exams, which were used to grade th 297 scripts examined in the study.

The 1998 ESL 3 Calibration paper reads as follows:

1. Content and Use of Appropriate Vocabulary • Does the student answer the question? • Does the student use relevant /appropriate vocabulary?	4 marks
2. Grammar and mechanics • Is appropriate grammar used? • Is there an absence of basic grammatical structures? • Is there a grasp of basic grammar? • Does the student use more sophisticated/complex forms? • Correct Use of capitalization, spelling and punctuation?	4 marks
3. Organization • Is there logical progression in the text? • Are there a clear introduction, development of ideas and a suitable conclusion?	2 marks

The calibration guideline used to grade the 1998 Spring exams was an ill-designed tool to measure student performance. Therefore, the design of the instrument could have influenced the results of teacher assessments of student papers, in addition to the other fundamental problems in the testing process.

The poor design of the instrument can be attributed to two main factors:

1. Criteria are not equally weighted. In category two, there are 5 established criteria to be met, but only 4 points allocated to the task. This indicates that a student has the potential to meet only 4 of the 5 criteria to get full marks.
2. The questions are not written as descriptors. Instead, many of the questions are "yes/no". This instrument implies in category one that if the teacher could answer yes to both questions, that full marks could potentially be awarded to the student's written script.

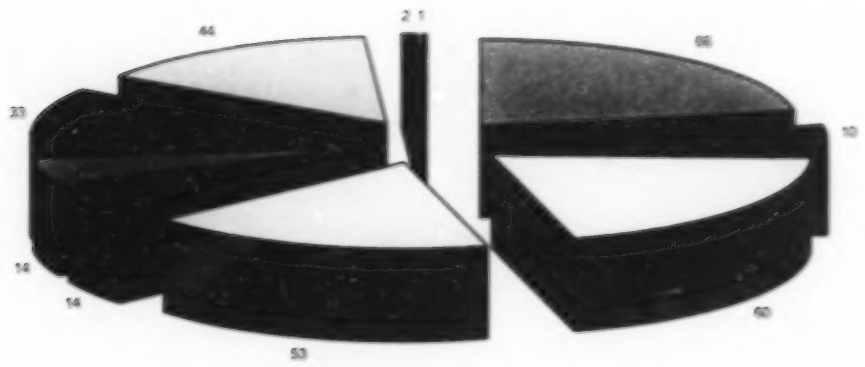
Sample scripts were used with teachers prior to grading to come to a consensus of what constituted a mark of "9" and what constituted a mark of "5". However, because of the ambiguity of the instrument, the lack of inter-rater reliability between markers/scores, the directive from the Unit Coordinator to eliminate the old practice of a third check to ensure consistency between all graders (that had been previously established by the past supervisor of testing) and the blatant disregard for calibration policy by some markers, resulted in inconsistent marking. The data collected from 297 random written exam scripts indicated 26.7% of the scripts were in some form of grading violation. The 88 papers with marking inconsistencies had the potential to

lead to unfair scoring practices (see figures 2a and 2b).

Figure 2a

1st and 2nd marker agreed with exact criteria *Note: Some packs, because blind marking was not used, had entire class scores with the exact same marks in every category.	66 papers	22%
1st and 2nd marker agreed with the score, but used different criteria in the assigned categories	10 papers	3.6%
Papers were off by 0.5 between the two markers	60 papers	20%
Papers were off by 1.0 points between the two markers	53 papers	17%
Papers were off by 1.5 points between the two markers	14 papers	4.7%
Papers were off by at least 2 or more points between the two markers	14 papers	4.7%
Papers in alleged violation : *No second marker's score is indicated on exam script box* *Papers scored without proper calculations* *Should have gone to arbitration* *No first or second marker score indicated on written script*	80 papers *33 papers* *44 papers* *2 papers* *1 paper*	26.7%

Figure2b.


44
2
1
66
33
10
14
14
60
53
Both Markers had the same calculations
Differential of .5 between the two calibrators scores
Differential of 1.5 between the two calibrators scores
Only one marker is noted on the paper
Differential of 3 pts or higher, but did not see an arbitrator
Both Markers had different calculations but the same score
Differential of 1 between the two calibrators scores
Differential of 2 between the two calibrators scores
Papers were scored but showed no records of calculations
No indication of a 1st or 2nd marker

5.0 Other Fundamental Grading Differences Between the HCT and the UGRU

The ACS of the HCT is responsible to provide reliable and valid examinations for the entire system. Statistical data, collected by Howell and Marsden (2000) and analyzed by in-house software packages, indicates the high level of accuracy in teacher banding when grading written scripts. This success can be attributed to:

- Teachers are given banding workshops as well as calibration sessions before marking.
- Banding is based on performance criteria found in the ESU Framework Scales.
- Scripts are not scored in class groups; instead, students are ranked alphabetically by college.
- The scripts are graded by a series of independent markers who are trained to use blind-marking practices.
- Each student and script are assigned a 4-digit code number. The code serves as an ID during the test.
- Teachers are given only 10 scripts to mark at a time by an exam official.
- Once the teacher has finished banding the writing scores, his/her pack is returned to a central administrator (CA). The CA rips off the results, which are then sent to a data entry room to be entered into the computer.
- A second marker grades the papers independently. Again when finished, his/her scores are torn off and sent to data entry. Because the first marker signs all the sheets with a special code number, it is impossible for the person to be the second or third grader.
- Only officials of ACS of HCT know the student codes and teacher codes.
- A third marker grades and turns in his/her banding scores for final data entry.
- After all three scores are entered into the software package, a double checker is used to ensure the scores have been entered correctly.
- Finally, the software decides the final band based on a combination of the 3 scores.

The software makes its determination by formulating at quarter band differentials. If the three banding scores are as follows, the final grade will be calculated as:

Marker 1	Marker 2	Marker 3	Marker 4	Final Result
4	4	4		4
4	3	4		4
4	3.5	3	*middle score	3.5
2	6	4	*Spread is too wide. ACS is alerted and a master banding expert is used to grade the exam	*Score is based on the expert banding decision.

- Upon completion of the data entry, the inter-marker data is reviewed by in-house designed software packages. Each individual teacher's performance is scored against his/her colleagues' grading. If a teacher is marking too high or too low, the software will alert ACS that there is a potential inconsistency.

The HCT's use of blind marking, code numbers, and alphabetized lists leads to:

- A clear policy that prevents teachers, students and superiors from pressuring the marker.
- A situation where no one marker is held accountable for a students' grade.
- A fair grading procedure for the student. No college-based bias or teacher bias.
- Random blind marking with no influence from other scores leads to a more equitable assessment.
- The demographics of script division help to keep and ensure fair testing practices.
- The quality of grading, due to blind marking, software intervention and banding criteria, can be benchmarked and scored against international standards.
- Criteria for grading is standardized and maintained year to year.
- Testing policies and procedures are maintained and reviewed yearly to further improve on the high standards of the institution.

The testing policies and practices up to Spring 2000 have been different in the UGRU Unit for the following reasons:

- Because teachers sign out packs, classroom teachers can see who is grading which set of papers.
- Teachers have the potential to pressure markers into being more or less difficult on a student when grading a paper.
- The 1st and 2nd marker is held accountable for the students' grades. If a paper is called into question, the classroom teacher can talk to an arbitrator who can change the mark accordingly. This system lacks inter-rater reliability.
- Students' names and id numbers are prominently displayed on exams. This practice has the potential to lead to gender bias. It also has the potential for teachers to make judgements about other teachers' abilities to teach and correct student writing.
- Because scripts are not separate from exams, this practice also allows the teacher to see other parts of the exam. Thus, letting the marker and classroom teacher know if they were weak in other areas and may potentially need the extra point in writing to help them "pass" the exam.
- During my seven-year tenure at UAEU the testing committee has faced many changes. The biggest challenge, in my opinion, is due to high turnover rates. For example, in my seven years, there have been 5 testing heads, 3 unit coordinators, 4 unit heads, and a multitude of curriculum heads and supervisors. None of these people had a university degree in the field of testing in hand. These manpower factors influence test writing and grading practices.

- Because standardized exams are not used at the end of levels, there are no international benchmarks.
- In-house exams are written to the student level and to cover material learned in a book. Again, proficiency is not measured by the IELTS, TOEIC, or PET in the UGRU.
- Writing criteria and calibration guidelines change year to year.
- The UGRU Unit has strict self-imposed censorship rules, unlike the HCT, which prevent certain topics, ideas, and concepts to be used or to come to light.
- Statistics are completed in-house by English teachers on the testing committee.
- In the past, classroom teachers were allowed to be the first or second markers on their own exams. However, this practice has changed. To the credit of the UGRU, teachers may only review the work of the first and second marker of their pack. They can no longer grade their own students' papers. If the classroom teachers disagree with the grades they may argue the papers through a random arbitration process.
- Second markers can see the scores awarded by the first markers at anytime during the marking process.
- Markers score class sets at one sitting.
- The transferring of grades to the computer from input sheets is based on the honesty system. A second checker signs his/her initials to verify the grades were transferred properly.
- There isn't a third party reviewer unless a paper is brought to arbitration. Scores are based on the first two markers who have to be within 1.5 points of one another.
- Despite having a calibration exercise an hour before marking, some teachers still ignore the guidelines and grade as they want to with disregard for the HCT policy. The UGRU does not have a software package, like the HCT, to alert people to unfair marking practices.
- Unlike the HCT, the UGRU does not have two external standardized tests for a student to pass in order to graduate.

6.0 **Future Directions for Study**

It would be recommended that the HCT continue on their current path of success and continue to further develop their already successful inter-rater reliability results on grading student written scripts.

Although the UGRU testing program has made consistent strides over the years on limited man-power and resources, some changes need to be made to further improve the foundations program. The following suggestions are only recommendations offered by the author who has worked in both systems. It would be advisable to have an outside testing consultant review the entire system using a top-down and bottom-up evaluation process for a more objective opinion. It would be equally advisable to have an internal evaluation from all faculty currently involved in teaching or testing courses in the UGRU.

It would be highly recommended that the UGRU Unit, hire a testing supervisor with a Ph.D. in testing and assessment. It would be suggested that the program consider restructuring the current testing program's marking practices. It would be recommended that the Unit Coordinator and assistant coordinators were not actively involved in testing.

It would be better if they could serve as a scanning team for inconsistencies, if software cannot be readily provided. It would be further recommended that standardized proficiency exams such as the PET and IELTS be implemented at the 2nd and 3rd levels of the EFL and ESL tracks as a course requirement to graduate from the foundation programs. It would be suggested that a banding system of descriptors be created and that consistent training by experts be offered to train teachers in independent blind marking to increase inter-rater reliability. It would also be advisable to not involve classroom teachers in the review or inputting of their students' exam papers. It would be advantageous to the Unit, to use more advanced statistical programs for data analysis and to use independent mathematicians to report on the validity of test questions and exam results. It would be advantageous to reinstate the previous policy of a third independent reviewer. A third marker grading a writing script should be incorporated and a fourth arbitrator should review papers where a spread of scores proves to be inconsistent. Exams should be independently reviewed to make sure that markers are grading to task. Lastly, Supervisors and teachers should be held more accountable for how they mark and score student papers.

It would be interesting to see if future writing scores would become more accurate and show more marking inter-rater reliability, if these suggestions were implemented. It would also be noteworthy, to see if the gender of the marker influenced scoring practices when correcting same sex and different sex students' papers.

**I would like to thank from the HCT Nicola Marsden, Elizabeth Howell, Dr. Christine Coombe for their endless support and help with this study and from the UGRU, Dr. Joesph Murphy, 1998 acting testing director and unit coordinator, for sanctioning the study and releasing the data. A special thanks to the help of Kathleen Bubniak and Manuela Zarifeh.*

Bibliography:

- Alderson, J and Wall, (1993) *Does Washback Exist?* Applied Linguistics 14 (2): 115-119
- HCT Homepage Mission Statement *ref:http://www.hct.ac.ae/general_info/frame.htm*
- HCT Aim Mission Statement *ref:http://imtcsamba.hct.ac.ae/gened/english/MissionStatement.html*
- HCT Curriculum Mission Statement *ref: http://imtcsamba.hct.ac.ae/gened/english/MissionStatement.html*
- HCT Assessment Mission Statement ref: *http://imtcsamba.hct.ac.ae/stand_eval/assessment/*
- Howell, E and Marsden, N (2000) interview notes from Wednesday September 13, 2000
- UGRU Mission Statement http://www.ugru.uaeu.ac.ae/
- UGRU 1998 Final Exam Calibration Guidelines, Testing Committee/ESL3 Curriculum Committee

Other Secondary References:

- Barlow L and Canning C (2000) "A Case Study of the Distance Learning Program & Testing at United Arab Emirates University", In Lynn Hendrichsen (ed.) Case Studies in International Distance Learning Programs, TESOL Publications in association with brigham Young University Press.
- Barlow, L and Canning, C (1998) "Strategies and Disruptions in Test Writing for UAE Distance Learning Students", ERIC Document, Microfische Documentation No. 426622

- Canning, C (1997) "Test Writing for UAE Distance Learning Students" In Christine Coombe (ed) Current Trends in English Language Testing, v2 October 1997/1998. Pp 58-72
- Coombe, C, Kinney J and C. Canning (1998) Issues in Evaluation of Academic Listening Tests, In Caroline Chapman and Diane Wall (eds) Language Testing Update (International Testing Journal), v 24m Autumn 1998.Pp 32-45
- Davison, T. (2000) *Why Do My Students Pass Tech and Fail English?* Higher Colleges of Technology Journal, 4:2 Pp 20-36
- Laihary, H. (2000) *Tutor Marked Assessments: How to Overcome the Problems in Setting and Marking Tests,* Higher Colleges of Technology Journal, 4:2 Pp 37-48
- Richardson, P (1999) *Assessment for Change at the Higher Colleges of Technology*, Higher Colleges of Technology Journal, 4:1 Pp 47-58
- Schley, N and Canning, C (1998) "Writing Effective Math and Computer Tests", EMCEE, 4:2 January 1998, Pp. 4-5

BEST COPY AVAILABLE